XP-002066935

01-09-1997
25.31

A candidate gene for familial Mediterranean fever

The French FMF Consortium

Familial Mediterranean fever (FMF) is an autosomal recessive disorder characterized by attacks of fever and serositis. In this paper, we define a minimal co-segregating region of 60 kb containing the FMF gene (*MEFV*) and identify four different transcript units within this region. One of these transcripts encodes a new protein (marenostrin) related to the ret-finger protein and to butyrophilin. Four conservative missense variations co-segregating with FMF have been found within the *MEFV* candidate gene in 85% of the carrier chromosomes. These variations, which cluster at the carboxy terminal domain of the protein, were not present in 308 control chromosomes, including 162 validated non-carriers. We therefore propose that the sequence alterations in the marenostrin protein are responsible for the FMF disease.

Familial Mediterranean fever (FMF; MIM 249100) is a recessively inherited disorder that primarily affects North African Jewish, Armenian, Turkish and Arab populations. The frequency of the disease gene in these populations is very high, with a carrier rate of 1:6 in North African Jews[1] and 1:7 in Armenians[2]. The disease is characterized by recurring attacks of inflammation in the peritoneum, synovium or pleura. Before the introduction of colchicine prophylaxis, the disorder was a major cause of amyloidosis with renal failure in FMF patients.

Because the biochemical defect is unknown, a positional cloning strategy was undertaken by several laboratories. The gene responsible for FMF was localized on the short arm of chromosome 16 in North African Jews, initially to a region of 9 cM (ref. 3) and subsequently to a region of about 4 cM between *D16S246* and *D16S523* (ref. 4). We further narrowed the FMF interval to a 250-kb genomic region between *D16S3070* and *D16S3275* by combining linkage disequilibrium and haplotype analysis[5]. This interval has been confirmed and refined to 200 kb[6]. We identified several founder haplotypes in various ethnic groups: haplotype S in North African Jews, haplotypes ARM1, ARM2 and ARM3 in Armenians, haplotype T in Turks, haplotypes ARA1 and ARA2 in Arabs from the Maghreb, and haplotype D in Druzes. We also showed that S, ARM1, T and ARA1 share a common origin, the Mediterranean haplotype MED.

A cosmid contig between *D16S3070* and *D16S3275* consisting of eight cosmids was entirely sequenced. A set of polymorphic markers that we identified in that interval was used to locate the disease gene in a 60-kb segment between *D16S2617* and *D16S3373*. Potential exons were identified in this 60-kb interval by exon amplification and computer analysis. Four functional genes were identified: a gene encoding an olfactory receptor, a gene encoding a zinc-finger protein, a gene encoding a product containing a ret-finger protein domain (marenostrin) and a gene tagged with expressed sequence tags (ESTs). Several missense mutations that co-segregate with FMF alleles were identified in the marenostrin-encoding gene—*MEFV*—and were not found on non-carrier chromosomes. Together with the location of this gene within the *MEFV* interval defined by genetic studies, these data indicate that the gene encoding marenostrin is identical to the *MEFV* locus. The name marenostrin derives from the Latin name of the Mediterranean sea, *mare nostrum*; it was chosen because most of the ethnic groups affected by FMF live around the Mediterranean basin.

Sequencing the *MEFV* candidate region

We first established a sequence-ready map of the *MEFV* candidate region that we had previously defined[5]. The candidate region was subcloned from YACs 16Gh7 and 633d12 (from the ICI and CEPH libraries, respectively) in the cosmid vector sCOGH2 (ref. 7). Cosmid contig assembly (Fig. 1) was based on STS content from genetic markers and cosmid and YAC insert ends. The contig continuity was checked by FISH on metaphase chromosomes and on extended fibres of genomic DNA. This indicated that the YAC 633d12 had a deletion of about 160 kb (Fig. 1). This part of the sequence was thus based on cosmid subclones derived from YAC clone 26Fe7 (250 kb long) from the ICI library, which contained markers *D16S3070* and *D16S3275*. Eight cosmids covered the candidate region except for a 3-kb segment (Fig. 1), which was cloned from a PCR product derived from amplification of genomic DNA. The total size of the *MEFV* candidate region was estimated to be 250 kb, between markers *D16S3070* and *D16S3275*.

A total sequence of 239 kb was assembled in five contigs separated by four sequencing gaps (between 200 and 650 bp), which were refractory to routine sequencing procedures. After fidelity assessment to the human sequence, this sequence was used to refine the candidate region and to define the embedded transcription units.

Group 1: Alain Bernot, Christian Clepet, Corinne Dasilva, Catherine Devaud, Jean-Louis Petit, Christophe Caloustian, Corinne Cruaud, Delphine Samson, Françoise Pulcini, Jean Weissenbach & Roland Heilig
Group 2: Cécile Notanicola, Cécile Domingo, Michaël Rozenbaum, Eldad Benchetrit, Rezzan Topaloglu, Marie Dewalle, Christiane Dross, Philippe Hadjari, Madeleine Dupont, Jacques Demaille & Isabelle Touitou
Group 3: Nizar Smaoui, Brigitte Nedelec, Jean-Philippe Méry, Habiba Chaabouni[1], Marc Delpech & Gilles Grateau

Group 1: Généthon, CNRS-URA1922, 1 rue de l'Internationale, 91000 Evry, France. Group 2: Laboratoire de Biochimie Génétique, Hôpital Arnaud de Villeneuve, 371 avenue du Doyen Giraud, 34295 Montpellier Cedex, France. Group 3: Biochimie et Biologie Moléculaire, Université Paris V, Institut Cochin de Génétique Moléculaire, Paris, France. [1]Service des maladies congénitales et héréditaires, Tunis, Tunisia.

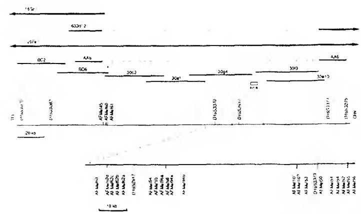

Fig. 1 Physical and genetic map of the *MEFV* candidate interval. Horizontal lines represent the extent of each clone (YAC clones: 16Gh7, 633d12 and 26Fe7, thick lines; cosmid clones AA6, BC2, BD6, 30c3, 30e10, 30g4, 30f9 and 30e10, thin lines). The deletions in the 633d12 YAC and in the AA6 cosmid are represented by dotted lines. Gaps in the sequence are represented by double slashes. AFMef numbers represent biallelic sequence variants except for AFMef45, which is a (CA)n microsatellite.

Refining the *MEFV* interval

In an analysis of historical recombinations in a Jewish founder haplotype, we had previously located *MEFV* to an interval of less than 250 kb between *D16S3070* and *D16S3275* (ref. 5). This region of interest has been confirmed and narrowed to about 200 kb, an interval framed by *D16S3082* and *D16S3373* (ref. 6). We recently showed haplotype-sharing among North African Jews, Armenians, Turks and Moslem Arabs (The French FMF Consortium, manuscript submitted). Examination of this common 'major Mediterranean haplotype' (MED) in our large, multi-ethnic panel of patients suggested a further refinement of the gene location between *D16S2617* and *D16S3373*. A search for sequence variations in the *MEFV* interval was undertaken to confirm the initial haplotype analysis and localize more accurately the putative ancestral crossover events that had taken place in the MED founder haplotype. Sequence analysis of the whole region indicated that there were no new microsatellites that could be used as genetic markers. We therefore looked for biallelic sequence polymorphisms. Several DNA fragments of 3 to 8 kb were sequenced and compared in patients and non-carrier individuals from FMF families, and 22 sequence variants were identified (Fig. 1).

All 22 markers were tested on FMF families; they show alleles in linkage disequilibrium with the disease locus, consistent with the common founder MED haplotype deduced from microsatellite analysis (The French FMF Consortium, manuscript submitted). The location of the *MEFV* interval could be refined as deduced from the analysis of founder haplotypes that were truncated by recombination events in two North African Jews (91-3 and 68-3) and two Armenians (A26-3 and A28-1) (Fig. 2). Patients 68-3 and 91-3 display one copy of the MED founder haplotype spanning the whole region and a second truncated copy that spans only the proximal part of this haplotype up to *D16S2617*. Whereas biallelic variants from the centromeric part are homozygous in both patients, sequence variants from the distal side of *D16S2617* are all heterozygous, revealing the presence of a region distinct from the founder haplotype in this segment. We concluded that patients 68-3 and 91-3 show historical recombinations allowing exclusion of the region distal to *D16S2617*. Although we could not exclude the loss of the mutation in the truncated founder haplotype segment, it is highly unlikely that this occurred twice, independently. In addition, this segment bears allele 18 for locus *D16S3275* in patient 68-3. This allele is in total disequilibrium with the disease.

Furthermore, patient A26-3 is homozygous for the MED haplotype in the portion distal to *D16S3373*. Similar to 68-3 and 91-3, heterozygous variants observed for proximal markers indicate the location of at least one historical recombination that excludes the region centromeric to *D16S3373*. The phase of *D16S3373*, afmef51, afmef52 and *D16S3275* could be determined in patient A26-3 and indicated that the founder alleles (5, G, AAA and 9, respectively) are on the same chromosome, which therefore carries the extended founder haplotype up to *D16S3275* (Fig. 2). Patient A28-1 shows only one copy of the MED haplotype spanning *D16S3124* to *D16S2617* and possibly *D16S3373*. The latter marker, however, does not show significant linkage disequilibrium (The French FMF Consortium, manuscript submitted). As markers afmef101 and afmef53 are homozygous for the non-founder allele (A/A and C/C, respectively) in this patient, these markers can be excluded from the MED founder haplotype carried by this chromosome (Fig. 2). Consequently, the historical breakpoint defined in A28-1 appears to be located distal to afmef101, further refining the *MEFV* centromeric boundary by a few more kilobases, from *D16S3373* to afmef101. We thus conclude that the *MEFV* region is flanked by *D16S2617* and afmef101-*D16S3373*, and according to the sequence map, spans a 60-kb interval (Fig. 1).

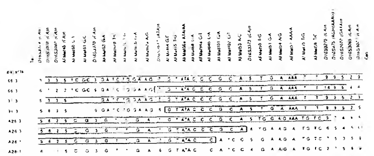

Fig. 2 Historical recombinations in the MED founder haplotype. FMF-carrier chromosomes found in North African Jews 68-3 and 91-3 and Armenians A26-3 and A28-1 are shown. The relative physical order of microsatellite and biallelic sequence markers is shown on the top. The extent of the Mediterranean founder haplotype is indicated by a frame, as shown previously (for *D16S3275* both allele 9 or 18 have been observed in this founder haplotype). S and ARM1 Mediterranean haplotypes differ for markers distal to *D16S3082*. *D16S3373* is the same locus as AFMef43, which we had previously characterized[5]; for biallelic markers, x/y indicate non-phased heterozygous markers; unavailable genotypings have been left blank. Tel, telomeric side; Cen, centromeric side.

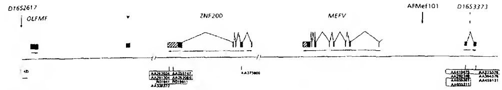

Fig. 3 Transcription map of the FMF region. The genomic structure of *ZNF200* and *MEFV* are indicated by diagrams showing exons (boxes) and introns (lines). Horizontal arrows indicate direction of transcription. The sequence is displayed with the centromeric direction towards the right. Locations of EST matches are indicated; when two ESTs represent each extremity of the same clone, they are boxed.

Sequence analysis of the *MEFV* candidate region

Candidate exons were identified in the genomic sequence spanning the *MEFV* interval using computer analysis. The sequence was first compared to public databases[8]. Comparison with EST databases[9,10] resulted in the detection of perfect matches with sixteen human ESTs (Fig. 3). The entire sequence was also analysed by three different exon-prediction programs: GRAIL[11], FGENEH[12] and GENIE[13]. A large number of sequences were rated as likely to be exonic. These predictions were tested as described in Methods and occasionally proved deceptive: we were able to detect transcription for only 20–35% of them (using FGENEH or GENIE, and GRAIL). However, regions predicted to be exonic by all three programs turned out to be true exons in almost all cases.

Exon predictions and comparisons with nucleic acid and protein databases detected four regions homologous to identified genes. The two most distal regions showed high similarity to genes encoding olfactory receptors[14], which are usually encoded by a single exon. A third region presented nucleic and amino acid similarities to zinc-finger coding regions[15] from a wide range of organisms. A fourth region showed similarities to several proteins containing an rfp-like domain (see below). A fifth area, identified by ESTs (Fig. 3), was found to lie outside the *MEFV* interval and could not be linked to other putative exons located within it. However, because several of these putative exons and regions of homology were probably part of larger transcription units, additional experimental validations and searches were undertaken with RT-PCR, PCR performed on cDNA libraries and RACE-PCR.

Gene identification in the *MEFV* interval

Exon trap analysis of cosmids 30f9, 30g4 and 30e10 led to the identification of 32 putative exons, including 19 located between *D16S2617* and *D16S3373*. Comparison of their sequence with the genomic sequence made it possible to align and orient these fragments along the genomic clones. The sequences identified by exon trapping were also compared to the EST databases and to the predicted exons. All these data were integrated to reconstruct tentative transcription units (Fig. 3).

We tried to confirm transcription of these putative exons and transcription units by PCR. Amplification products from human cDNA libraries of sizes compatible with the predictions were sequenced. The cDNAs from the two multiple-exon genes were extended with RACE-PCR. We deduced the positions of the splice junctions from alignment of the putative exons with the genomic sequence; all of them displayed *bona fide* 5′ and 3′ consensus motifs.

cDNA fragments corresponding to the *ZNF200* gene were amplified from a liver cDNA library and assembled into a 3,060-nucleotide transcript. This gene is composed of five exons distributed over 14 kb of genomic DNA. Several ESTs are included in the 3′-most exon of *ZNF200*, and sequence AA375800 covers the second exon of this gene (Fig. 3).

A 1.9 kb cDNA sequence encoding marenostrin was obtained from eight overlapping cDNA sequences amplified from a leukocyte cDNA library. The gene encoding marenostrin contains at least ten exons spread over more than 10 kb of genomic DNA (Fig. 3). No EST matches were found in any public sequence data banks. When northern blots of poly(A)$^+$ RNAs from various human adult tissues were hybridized to a cDNA probe corresponding to this gene, a faint band, of approximately 4 kb, was detected in mRNA from peripheral blood leukocytes (PBLs). No expression was detected in the other tissues tested (spleen, thymus, prostate, testis, uterus, small intestine, colon, stomach, thyroid, spinal cord, lymph node, trachea, adrenal gland, bone marrow, heart, brain, placenta, lung, liver, muscle, kidney and pancreas). Preliminary screening of a PBL cDNA library of 1,200,000 clones using the 1.9-kb 3′ probe yielded ten positive clones. However, expression of this gene was detected in spleen and leukocyte cDNA libraries by PCR and by use of RT-PCR *in synovia*. The PCR products were of the expected sizes, and their sequence was identical to that of fragments of the gene encoding marenostrin.

Analysis of the coding sequences

Of the two regions homologous to olfactory receptor genes, one encodes an uninterrupted open reading frame (ORF), which is likely to correspond to a functional gene. We call this putative gene *OLFMF*, although the approved gene symbol is currently under review by the HUGO nomenclature committee. The conceptual translation of the second region homologous to this superfamily does not contain a methionine at its amino terminal end, and as all other known olfactory receptors are encoded by single exons, we concluded that this ORF is likely to be a pseudogene (Fig. 3).

The *ZNF200* gene contains a single ORF of 394 amino acids. This amino-acid sequence corresponds to a polypeptide with no evidence of a transmembrane domain. The last exon encodes five contiguous zinc-finger domains[15]. The first two of them fit perfectly the consensus sequence of the C2H2 group (Y/F-X-C-X2-4-C-X3-F-X5-L-X2-H-X3-4-H-X6-7), and the H-C link following each of them is also strictly conserved (TGE-R/K-P-F/Y-X). These two domains are followed by three other zinc-finger domains for which the structure and link are less conserved. This gene shares homologies with a large number of zinc-finger–containing genes.

The *MEFV* cDNA contains a single ORF of 477 amino acids extending from nucleotide 1 to 1433 (Fig. 4*a*). This sequence does not include the initiator methionine codon and thus appears to lack the 5′ end of the corresponding mRNA. The carboxy-terminus of this polypeptide chain (from residue 83 to the end) exhibits significant similarities with the rfp-domain, the prototype of which was first described in the RET finger protein (RFP)[16]. It has since been observed in the butyrophylin precursor[17], a zinc-finger DNA-binding protein from the amphibian *Pleurodeles waltii*[18], the mouse RPT-1 protein[19] and the Sjögren syndrome type-A antigen[20]. The RFP-like region is represented in ProDom by four domains[21] (4240, 7680, 9608, 7736; Fig. 4*b*). No three-dimensional structure is currently available in protein databases for any of them. The amino-terminal part of the conceptual translation (residues 1–82) is a Cys/His-rich region that may function in metal binding.

a

b

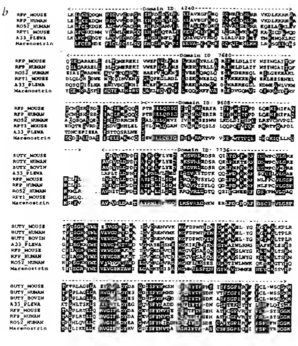

Fig. 4 a, Partial amino-acid sequence of marenostrin as deduced from its cDNA sequence. **b**, Comparison of the conceptual translation to the domains defined in ProDom: domain 4240 (residues 81–138 of marenostrin), domain 7680 (139–219), domain 9608 (220–263), domain 7736 (293–461). RFP_MOUSE and RFP_HUMAN, ret finger proteins from mouse and human; RO52_HUMAN, Sjogren syndrome type A antigen; RPT, Regulatory protein, T lymphocyte, 1; A33_PLEWA, zinc finger DNA-binding protein from *Pleurodeles waltii*; BUTY_MOUSE, BUTY_HUMAN, BUTY_BOVIN, Butyrophilin from mouse, human, and cow. Dashes mark gaps introduced to allow optimal alignment. Areas of homology are highlighted.

Sequence variations in FMF patients

Because no gross rearrangement of genomic DNA was detected in FMF patients with long-range PCR (data not shown), the coding exons from the genes encoding OLFMF, ZNF200 and marenostrin were sequenced in seven FMF patients who were homozygous for one of the founder haplotypes and in eleven unaffected controls (from various populations of European descent). The sequencing was carried out on genomic PCR products to analyse the entire coding region of these genes.

The olfactory receptor exhibited a non-conservative substitution in the Arabian individuals we tested, but this substitution was also found in unaffected or control individuals from various ethnic groups. The *ZNF200* gene did not show any sequence variation in the coding exons.

Four different sequence alterations that introduced changes in *MEFV* were identified in the 3′-most exon in the different ethnic groups (Table 1). The first change (Med variation) consisted of an A→G transition that changed a methionine codon (ATG) into a valine codon (GTG). It was observed in families of different ethnic origins (Jewish, Armenian, Turkish and Arabian) sharing a MED haplotype. Second, a G→A transition (Ara2 variation) that changed the same methionine codon into an isoleucine codon (ATA) was observed in an Arabian family bearing an ARA2 haplotype. A third alteration (Arm3 variation) was a T→C transition that changed a valine codon (GTT) into an alanine codon (GCT). It was observed in patients with a Druze haplotype (D) and in patients with an ARM3 haplotype. Finally, a change of another methionine into isoleucine (G→C transversion; Arm2 variation) was found in Turkish and Armenian patients with an ARM2 haplotype. Other nucleotide substitutions were identified in other exons, but these resulted in synonymous codons (data not shown).

Table 1 • Mutations in *MEFV* in FMF families

Haplotype	Nucleotide change	Position	Coding effect
Med	ATG→GTG	1170	Met → Val
Ara2	ATG→ATA	1172	Met → Ile
D	GTT→GCT	1267	Val → Ala
Arm3	GTT→GCT	1267	Val → Ala
Arm2	ATG→ATC	1130	Met → Ile

Analysis of control individuals

A number of control analyses were performed on genomic DNAs with the aid of the amplification refractory mutation system (ARMS[22], for the Med and Ara2 variants) or, when applicable, a combination of appropriate PCR primer sets and restriction enzymes (for the Arm2 and Arm3 variants).

We first explored the possibility that the observed sequence variants could correlate with candidate mutations, and looked for co-segregation of the marenostrin sequence variants and the disease. We therefore screened both carriers and non-carriers from affected kindreds. Among all pedigrees tested (more than forty), one of the four marenostrin sequence variants was detected in all obligate carriers.

The variations were then examined in a large number of carrier chromosomes, and each alteration was found to be restricted to a single founder haplotype or on chromosomes that were tentatively linked to one of the founder haplotypes (Fig. 5). The Med (Met→Val) variant is found in 83% of the carrier chromosomes of North African Jews, in strict correlation with all MED haplotypes. In other ethnic groups, several carrier chromosomes, which displayed allelic combinations resembling the founder haplotypes, did not contain the founder amino-acid variants (Fig. 5). In the latter cases, however, the relevance to the founder haplotype is more questionable, as they usually lack one or more of the flanking microsatellites showing the highest linkage disequilibrium.

To exclude the possibility that the variants were simply common polymorphisms in the population, we screened a set of control DNA samples for each variation. The sequence variants were absent from a panel of DNA from 73 unrelated parents or grandparents (146 chromosomes) of the CEPH families from various groups of European descent[23].

Given the elevated frequency of the mutation in North African Jews and Armenians (8% and 7%, respectively), we restricted our negative controls to the non-carrier chromosome of FMF carriers in

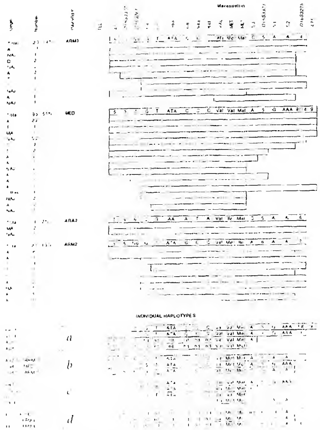

Fig. 5 Founder FMF haplotypes observed in different affected populations. Haplotypes are grouped to schematize the chromosome segment retained from each of the four founder haplotypes associated with a marenostrin sequence variation. Number refers to the number of carrier chromosomes bearing the framed allelic combination. The percentages represent the fractions of carrier chromosomes with the indicated haplotype. 24 of the carrier chromosomes analysed (14%) did not bear one of the four founder missense mutations. Individual haplotypes presented at the bottom of the figure: *a*, Recombinant MED chromosomes from Fig. 2. *b*, Carrier chromosomes with founder-like haplotypes for which no mutation has yet been found. *c*, Carrier chromosomes with no evident founder haplotype but with a founder mutation. *d*, Non-carrier chromosomes with founder-like haplotype but no founder mutations. Ethnic origins: A, Armenians; NAJ, Non-Ashkenazic Jews (including North African and Iraqi Jews); D, Druzes; T, Turks; MA, Arabs from the Maghreb; O, others.

these populations. None of the four Med, Ara2, Arm2 and Arm3 alterations were observed in any of the 162 non-carrier chromosomes studied. In particular, non-carrier chromosomes showing similarities to the founder haplotypes did not carry these variations (Fig. 5). The four alterations observed are thus strictly correlated with the carrier haplotypes.

Discussion

To identify the gene responsible for FMF, we sequenced the entire 250-kb candidate region. A minimal contig, consisting of eight cosmid clones and one 3-kb PCR fragment, was sequenced. The frequent rearrangements occurring in YACs made it necessary to ensure that the sequence obtained accurately represented human genomic DNA. A 160-kb block that was deleted in YAC 633d12 was then subcloned from another YAC. A subdomain of this block (also underrepresented in cosmid clones) included the 3-kb PCR fragment.

The extensive polymorphism analysis that we carried out using newly developed biallelic sequence variants from the *D16S3070–D16S3275* interval resulted in an accumulation of confirmations of the common MED founder effect that we observed previously among Armenians, Turks, Jews and Arabs. These markers also proved useful for confirming putative ancestral crossovers that had taken place in the MED founder haplotype, and helped narrow the *MEFV* locus to a 60-kb minimal region flanked by *D16S2617* and AFMet101–*D16S3373*.

The transcriptional units identified in the critical 60-kb *MEFV* interval were based on analyses of the genomic sequence and exon-trapping experiments. We chose these methods because they are genome-based and thus expression independent: exon trapping takes advantage of the splicing mechanism in living cells, and sequence analysis relies mainly on the use of computer programs. Validation of the predicted exons led to the characterization of three genes: one encodes an olfactory receptor, the second encodes a protein showing similarities with zinc-finger proteins and the third encodes marenostrin, a new member of a family including RFP[17], BT[1], Pwa33 (ref. 18), RPT-1 (ref. 19) and SS-A[17]. A fourth gene was only characterized by its alignment with ESTs. It is unlikely that additional genes remain unidentified in this interval.

We analysed the first three genes for sequence variation by comparing seven patients homozygous for one of the founder haplotypes to eleven control samples. We sequenced genomic PCR products to analyse the known coding region of these genes and identified missense mutations in three different codons of the last exon of the gene encoding marenostrin. Several lines of evidence support the implication of this gene in the pathogenesis of FMF. First, there is a perfect correlation between the affected phenotypes and the DNA alterations. Second, we found none of the sequence modifications in either non-carrier or other control individuals including non-carrier chromosomes showing similarities to one of the founder haplotypes. Third, none of the other sequence variations that we observed in *MEFV* or another gene contained within the 60-kb FMF interval were associated with disease.

The occurrence of the same sequence alteration (ATG-GTG; Met/Val) in families bearing one of the MED haplotypes (S, ARM1, T or ARA1) indicates that this event is ancient, and confirms the founder effect at the origin of a large fraction of FMF cases from the Mediterranean basin. The presence of a single variation common to both the ARM3 and Druze haplotypes that share surrounding alleles also indicates that these chromosomes share a common core haplotype. Conversely, the occurrence of the other distinct mutations in other founder haplotypes confirms the different origins of the corresponding chromosomes.

The definition of the haplotype is reinforced by an analysis using the four marenostrin amino-acid variants. Each variation is in complete disequilibrium with a founder haplotype. However,

when the haplotype definition is less stringent and does not include the multiallelic polymorphisms in high linkage disequilibrium, the link with the founder haplotype remains tentative. Alternatively, the presence of one of the four sequence variants in such cases could result from recurrent mutations that created additional carrier chromosomes. Under this hypothesis, the fact that identical independent variations are associated with the disease would serve to demonstrate their deleterious effect.

It is difficult to speculate about the function of marenostrin by comparing it with related proteins. A nuclear localization and nucleic acid–binding properties or a role in regulating gene expression has been proposed for SSA-1, Pwa33, RPT-1 and RFP, which also contain amino-terminal zinc-finger motifs. Thus, marenostrin could belong to a family of nucleic acid–binding regulatory factors, and may regulate gene expression. The rpt-1 protein has been implicated in the regulation of the alpha chain of the interleukin-2 receptor; members of this gene family can thus be involved in the regulation of expression of immune-related proteins, an observation that could be related to the inflammation observed in FMF. Moreover, marenostrin is expressed in leukocytes, which are engaged in both immune and inflammatory responses. On the other hand, it should be noted that butyrophylin, which also contains an rfp-like domain but lacks zinc fingers, is a transmembrane glycoprotein expressed in mammary tissue during lactation. Thus, additional insight into the function of marenostrin awaits the complete cloning and sequencing of the rest of the transcript, which is in progress.

All the base alterations described here resulted in a conservative change of a hydrophobic amino acid. Although this kind of amino-acid change often has little or no phenotypic effect, its impact can be much more dramatic. For instance, the most prevalent mutations in the gene encoding transthyretin, which results in amyloid polyneuropathy, are Val→Met, Leu→Met and Val→Ile substitutions[24]. These mutations produce amyloidosis (although in this case it is a primary syndrome—in contrast to FMF, where the amyloidosis is probably secondary). Several mechanisms—such as modification of conformation or stability, alteration of a binding site or other sites of interaction—could account for phenotypic effects of such mutations. It is also striking that these four sequence variants from the carrier chromosomes are clustered in the same exon. It could be argued that these changes are in strict linkage disequilibrium with the mutations but are not the mutations themselves. It is, however, extremely unlikely that four such variations accumulated randomly in a very limited stretch of DNA sequence in carrier chromosomes only.

The discovery of this novel gene may have important clinical implications; 72% of the patients in our sample could be characterized by one or two of four mutations. This should enable clinical geneticists to offer reliable diagnostic tests for this ill-diagnosed disease, for which effective therapy is available.

Methods

Cosmid contig construction. YACs 16Gh7, 633d12 and 26Fe7 were partly digested with *Sau*3AI, dephosphorylated, ligated in the *Bam*HI site of cosmid vector sCOGH2 (ref. 7) or RH3 (R.H., unpublished), *in vitro* packaged and used to transfect the *Escherichia coli* strain STBL2, according to standard protocols. Cosmids from clones hybridizing to human DNA were purified and spotted with a Hybaid robot onto nylon membranes (Hybond N+). Filters were hybridized to non-radioactive probes synthesized from STS primers and cosmid ends.

Primers were labelled with biotin-16-ddUTP (Boehringer Mannheim) using terminal transferase and hybridized to filters according to standard protocols. Insert-end probes were generated by asymmetric PCR based on one vector primer (hGHlink.up: 5′–ACAATGGGAGCTGGTCTC–3′ or hGHlink.lo: 5′–TAGGAGAAGGACCGCCCA–3′) in a 50-µl reaction volume, containing 500 ng cosmid DNA, 10 µM digoxygenin-11-dUTP (Boehringer Mannheim), 90 µM dTTP, 100 µM each d(ACG)TP, 0.5 µM primer, 10 mM Tris HCl, pH 8.8, 50 mM KCl, 1.5 mM $MgCl_2$ and 0.1% Triton X-100. *Taq* DNA polymerase (1 U, Cetus) was added after 5 min at 94 °C, and the mixture was subjected to 30 cycles as follows: 25 s at 94 °C, 30 s at 60 °C and 1 min at 72 °C. Probes were verified on a 0.8% agarose gel and purified on G25 Sephadex spin columns. Filter hybridizations were performed in the DIG Easy Hyb buffer (Boehringer Mannheim), with a probe pre-hybridized with Cot-1 DNA. For signal detection, the CDP-Star TM substrate of the alkaline phosphatase was used (Boehringer Mannheim). Data analysis and contig assembly were performed with the Contig package[25].

Pairwise end sequencing[26]. For each cosmid, fragments from partial digestion by *Cvi*JI were gel-sized for a range of 8–10 kb, subcloned in plasmid pBC-SK+ (CmR, Stratagene) and sequenced on both ends. Assembly of a mean of 300 exploitable reads, using PHRED and PHRAP software (P. Green, pers. comm.), resulted in a scaffold of subclones for each cosmid, fixing the sequence contigs (in terms of distance and relative orientation) as well as the resulting gaps (in position and size), thus monitoring the finishing operations, by primer walking for example. This procedure was found to be useful for resolution of highly repetitive regions or nearly exact duplications.

Contig and sequence validation. Chimaerism of YACs was checked by fluorescence *in situ* hybridization (FISH) on metaphase spreads from healthy donors, using total YAC clone DNA or inter-Alu PCR products as probes[27]. Fibre FISH mapping was performed by co-hybridization of some selected cosmids as probes (biotin-16-dUTP or digoxygenin-11-dUTP labelled) on extended DNA[28] prepared from (i) genomic lymphoblastoid cell lines from healthy donors, (ii) the non-chimaeric YAC 26Fe7 as reference and (iii) the deleted YAC clone 633d12 (E.P., in preparation). Biotin and digoxygenin-labelled probes were detected by Texas Red and FITC, respectively, and simultaneously visualized with a double band pass filter using a Leica DMRB epifluorescent microscope; images were recorded with a CCD camera. Successive rounds of co-hybridization of three cosmids ensured coverage of the YAC 633d12 deletion.

Fidelity to the human sequence was assessed (i) by comparing the restriction map deduced from the sequence (for *Eco*RI, *Bgl*II, *Not*I, *Eag*I, *Mlu*I and *Bss*HII) to the restriction patterns observed on all the assembled cosmids in the region of interest or to published data[9]; (ii) by systematic long-range PCR analysis of control DNAs; and (iii), in some critical instances, by Southern blot analysis, using digested DNAs from control samples. Furthermore, all detected exons and more than half of the sequence of the critical region (more than 35 kb in total) were resequenced from 'normal' and 'affected' DNAs.

Development of biallelic sequence markers. Based on sequencing data, eight PCR fragments (3–8 kb long) were designed in the *MEFV* interval. LA PCR (Takara) was used to prepare the corresponding DNA fragments from patients and healthy individuals from FMF families. These PCR templates were then purified through Microcon devices (Amicon) and subjected to direct dye-terminator cycle sequencing. Sequence comparisons were performed with the Staden 96 [illegible] package. Primers were designed with the Oligo 4.0 software[29]. Detailed information on new markers is available from the Genome Data Base (GDB).

Genetic markers. Information on microsatellite markers was from GDB. AFMe151 and AFMe152 have been described elsewhere (The French FMF Consortium, manuscript submitted).

Gene prediction. The BLAST[30] programs were used to determine similarities and identities to known genes using a non-redundant compilation of the EMBL and GenBank databases. Amino acid comparisons were performed by translating DNA sequences into all six potential reading frames and comparing translations to protein sequences in a non-redundant Swiss Prot and PIR database using the program BLASTX. Protein domain homologies were found by searching the ProDom protein database.

Exon predictions were performed using several programs. The GRAIL program was accessed through the GRAIL e-mail server, at GRAIL@ornl.gov. Sequences were sent with the option -2. BCM GeneFinder was accessed at service@bchs.uh.edu, with the keyword 'fexh' in the subject line of the message. GENIE was accessed through the GENIE server at http://www-hgc.lbl.gov.

Exon trapping. Cosmid DNA was partly digested with *Sau*3AI and ligated into the pSPL3 splicing vector (BRL). The ligation reaction was transformed into an *E. coli* XL1 Blue host by electroporation. Plasmid DNA was recovered and electroporated into COS-7 cells. RNA was isolated 48 h after transfection and converted into single-stranded cDNA with the SuperScript II (BRL) reverse transcriptase. Single-stranded cDNA was converted to double-stranded DNA through six PCR amplification cycles. To eliminate vector-only and false-positive products, *Bst*XI was added to the reaction and incubated overnight. RNA/PCR products were then cloned into the pAMP vector (BRL) and individual clones were sequenced.

cDNA amplification. cDNA fragments corresponding to the marenostrin and ZNF4II proteins were amplified from cDNA libraries or from Quick-Clone cDNA (Clontech). Primers were designed from the exon trapped product sequences, and used in PCR in standard procedures. 5′ and 3′ RACE was performed to characterize the extremities of these genes using the Marathon cDNA amplification kit (Clontech) according to the manufacturer's instructions. Nested PCRs were carried out using oligonucleotides from the adaptator (AP1 and AP2) and internal oligonucleotides from the cDNA sequences. PCR products were cloned into the TA vector, and recombinant clones sequenced.

Mutation identification. The exons and flanking introns were amplified with PCR and specific primers. For *MEFV*, the primers were as follows: [illegible]

PCR conditions were denatured at 95 °C for 10 min, 34 cycles at 95 °C for [illegible], [illegible] °C for 30 s and 72 °C for 3 min, and final extension at 72 °C for [illegible] min. PCR products were purified with Sephadex P100 chromatography and sequenced directly using specific primers and AmpliTaq FS Dye terminator cycle sequencing kit and ABI Prism 377 DNA sequencer.

Screening of mutations

[illegible] variants. The [illegible] most exon was amplified from genomic DNA with the primers p12.2 (5′-[illegible]-3′) and p10.1 (5′-CTCCCGTACTTCCTCCTC[illegible]-3′). PCR products were digested as suggested by the manufacturer by *Mn*I and *Hin*fI for the screening of T/C (1267) and G/C (1130) mutations, respectively. In the first case, an *Mn*I site is created, and a *Hin*fI site is destroyed in the second case. Products of digestion were analysed on 6% polyacrylamide gels.

Med and Ara2 variants. The A/G (1170) and G/A (1172) mutations do not alter any known restriction site, and so allele-specific primers were devised. The Med mutation was distinguished with the met1 (5′-TGGTACTCATTTTCCTTCAT-3′) and p12.2 (above) oligonucleotides, and with val1 (5′-TGGTACTCATTTTCCTTCAC-3′) and p12.2. The Ara2 mutation was monitored with met2 (5′-TGGTACTCATTTTCCTTC-3′) and p12.2 and with del (5′-CTGGTACTCATTTTCCTTT-3′) and p12.2. PCR products were analysed on 3% NuSieve 1% Seakem gels.

Northern-blot analysis. Northern blots (Clontech) were prehybridized in ExpressHyb buffer (Clontech) and hybridized in the same solution containing a ^{32}P-labelled random-primed cDNA probe. The membranes were washed at 50 °C with 0.1× standard saline citrate (SSC) and 0.1% SDS, then exposed to X-ray film at −80 °C for 7 days.

EMBL accession numbers. *ZNF200*, Y14443; *OLFMF*, Y14442; marenostrin, Y14441.

Acknowledgements

We thank the MEF families who have participated in this study as well as the physicians who helped in the initial patient collection: D. Cattan, J. Medaxian, [illegible], H. Ben Mâlik, [illegible], K. Ozen and R. M[illegible]. We also thank C. Lizarnes and L. Brais [illegible] with computer analysis, E. Wunderle and N. Bernex for technical assistance and I. Hazan and S. [illegible] for critical reading of the manuscript. This work was supported by the Association Française contre les Myopathies (AFM).

Received 30 July; accepted [illegible] August 1997.

1. Daniels, M., Shohat, T., Brenner-Ullman, A. & Shohat, M. Familial Mediterranean fever: high gene frequency among the non-Ashkenazic and Ashkenazic Jewish populations in Israel. *Am. J. Med. Genet.* **55**, 311–314 (1995).
2. [illegible] Familial Mediterranean fever in Armenians [illegible] high gene frequency. *Am. J. Med. Genet.* **34**, 168–172 (1989).
3. Aksentijevich, I. *et al.* Refined mapping of the gene causing familial Mediterranean fever by linkage and homozygosity studies. *Am. J. Hum. Genet.* **53**, 451–461 (1993).
4. Levy, E. *et al.* Linkage disequilibrium mapping places the gene causing familial Mediterranean fever close to *D16S246*. *Am. J. Hum. Genet.* **58**, 523–534 (1996).
5. The French FMF Consortium. Localization of the familial Mediterranean fever gene (FMF) to a 250-kb interval in non-Ashkenazi Jewish founder haplotypes. *Am. J. Hum. Genet.* **59**, 603–612 (1996).
6. Sood, R. *et al.* Construction of a 1-Mb restriction-mapped cosmid contig containing the candidate region for the familial Mediterranean fever locus (MEFV) on [illegible] **42**, [illegible] (1997).
7. Datson, N. *et al.* Scanning for genes in large genomic regions: cosmid-based exon trapping of multiple exons in a single product. *Nucleic Acids Res.* **24**, 1105–1111 (1996).
8. Altschul, S. *et al.* Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
9. Adams, M.D. *et al.* Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature* **377**, 3S–174S (1995).
10. Hillier, L. *et al.* Generation and analysis of 280,000 human sequence tags. *Genome Res.* **6**, 807–828 (1996).
11. Xu, Y., Mural, R., Shah, M. & Uberbacher, E. Recognizing exons in genomic sequence using GRAIL II. in *Genetic Engineering: Principles and Methods* Vol. 16 (ed. Setlow, J.) 241–253 (Plenum, New York, 1994).
12. Solovyev, V., Salamov, A. & Lawrence, C. Predicting internal exons by oligonucleotide composition and discriminant analysis of spliceable open reading frames. *Nucleic Acids Res.* **22**, 5156–5163 (1994).
13. Kulp, D., Haussler, D., Reese, M. & Eeckman, F. A generalized hidden Markov model for the recognition of human genes in DNA. in *ISBM 96* (ed. AAAI) 134–142 (MIT Press, St. Louis, Missouri, 1996).
14. Buck, L. & Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**, 175–187 (1991).
15. Klug, A. & Rhodes, D. Zinc fingers: a novel protein motif for nucleic acid recognition. *Trends Biochem. Sci.* **12**, 464–467 (1987).
16. Takahashi, M. & Cooper, G. Ret transforming gene encodes a fusion protein homologous to tyrosine kinases. *Mol. Cell. Biol.* **7**, 1378–1385 (1987).
17. Jack, L. & Mather, I. Cloning and analysis of cDNA encoding bovine butyrophilin, an apical glycoprotein expressed in mammary tissue and secreted in association with the milk fat globule membrane during lactation. *J. Biol. Chem.* **265**, 14481–14486 (1990).
18. Bellini, M., Lacroix, J. & Gall, J. A putative zinc-binding protein on lampbrush chromosome loops. *EMBO J.* **12**, 107–114 (1993).
19. Patarca, R. *et al.* rpt-1, an intracellular protein from helper/inducer T cells that regulates gene expression of interleukin 2 receptor and human immunodeficiency virus type I. *Proc. Natl. Acad. Sci. USA* **85**, 2733–2737 (1988).
20. Tsugu, H., Horowitz, R., Gibson, N. & Frank, M. The location of a disease-associated polymorphism and genomic structure of the human 52-kDa Ro/SSA locus (SSA1). *Genomics* **24**, 541–548 (1994).
21. Gouzy, J., Corpet, F. & Kahn, D. Graphical interface for ProDom domain families. *Trends Biochem.* **21**, 493 (1996).
22. Newton, C. *et al.* Analysis of any point mutation in DNA: the amplification refractory mutation system (ARMS). *Nucleic Acids Res.* **17**, 2503–2516 (1989).
23. Dausset, J. *et al.* Program description: Centre d'Etude du Polymorphisme Humain (CEPH): Collaborative genetic mapping of the human genome. *Genomics* **6**, 575–578 (1990).
24. Almeida, M. *et al.* Haplotype analysis of common transthyretin mutations. *Hum. Genet.* **96**, 350–354 (1995).
25. Mott, R., Grigoriev, A., Maier, E., Hoheisel, J. & Lehrach, H. Algorithms and software tools for ordering clones libraries: application to the mapping of the genome of *Schizosaccharomyces pombe*. *Nucleic Acids Res.* **21**, 1965–1974 (1993).
26. Coquelle, A., Pulcini, F., Danglot, G. & Fauvet, D. Mapping of 22 YACs on human chromosomes by FISH using yeast DNA Alu PCR products for competition. *Ann. Genet.* **39**, 64–68 (1996).
27. Roach, J., Boysen, C. & Hood, L. Pairwise end sequencing: a unified approach to genomic mapping and sequencing. *Genomics* **26**, 345–353 (1995).
28. Rosenberg, C. *et al.* High resolution DNA Fiber-FISH on yeast artificial chromosomes: direct visualization of DNA replication. *Nature Genet.* **10**, 477–479 (1995).
29. Wiegant, J. *et al.* High-resolution *in situ* hybridization using DNA halo preparations. *Hum. Mol. Genet.* **1**, 587–591 (1992).
30. Bonfield, J.K., Smith, K.F. & Staden, R. A new DNA sequence assembly program. *Nucleic Acids Res.* **23**, 4992–4999 (1995).
31. Rychlik, W. & Rhoads, R. A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and *in vitro* amplification of DNA. *Nucleic Acids Res.* **17**, 8543–8551 (1989).